A Recent Discovery in Anomaly Detection

Don Hush, Patrick Kelly, Ingo Steinwart, Clint Scovel, CCS-3

Anomalies are often described as rare or unusual events. This notion can be represented mathematically by defining anomalies to be points with low probability density value. Detecting anomalies in practice is hard because we don't know the probability density function, and therefore must design a detector based on incomplete information. In particular, we are given empirical information consisting of a finite collection of points sampled from the probability density function and then asked to design a detector that makes predictions (guesses) about which points are anomalous. The fraction of incorrect predictions (on both current and future points) is called the error rate, and our goal is to design a detector whose error rate is as small as possible.

One of the most fundamental and longstanding issues in anomaly detection is that there is no way to obtain a reliable estimate of the error rate for a detector, regardless of how it is designed. This means that there has been no way to quantitatively compare two (or more) detectors to determine which one is best. Researchers in CCS-3 have made a recent discovery that resolves this situation. In particular they have discovered a performance measure that is calibrated with respect to the error rate, which means that it can be used as a replacement for the error rate. Furthermore, this new performance measure can be reliably estimated in practice. Consequently it is now possible to make quantitative evaluations of performance, and to objectively answer the question, "Which of these detectors is more accurately identifying the anomalous points?" Furthermore, this discovery has opened the door to a whole new class of design methods that work by optimizing this new performance measure. In particular, CCS-3 researchers have invented a new design method that is guaranteed to be computationally efficient and to provide near-optimal results for nearly all probability densities encountered in practice.

For further information contact Don Hush at dhush@lanl.gov.

References:

D. Hush and J. Howse, "Anomaly Detection on Graphs," Los Alamos National Laboratory Technical Report LA-UR-05-8440 (2005).

D. Hush, P. Kelly, C. Scovel and I. Steinwart, "Provably Fast Algorithms for Anomaly Detection," Los Alamos National Laboratory Technical Report LA-UR-05-4367 (2005).

I. Steinwart, D. Hush, and C. Scovel, *Journal of Machine Learning Research,* **6,** 211-232 (2005).

C. Scovel, D. Hush, C. Scovel and I. Steinwart, *Analysis and Applications,* **3**(4), 356-371 (2005).

I. Steinwart, D. Hush, and C. Scovel, *Neural Information Processing Systems,* **17,** 1337-1344 (2005).

Funding Acknowledgments

- Los Alamos National Laboratory Directed Research and Development Program

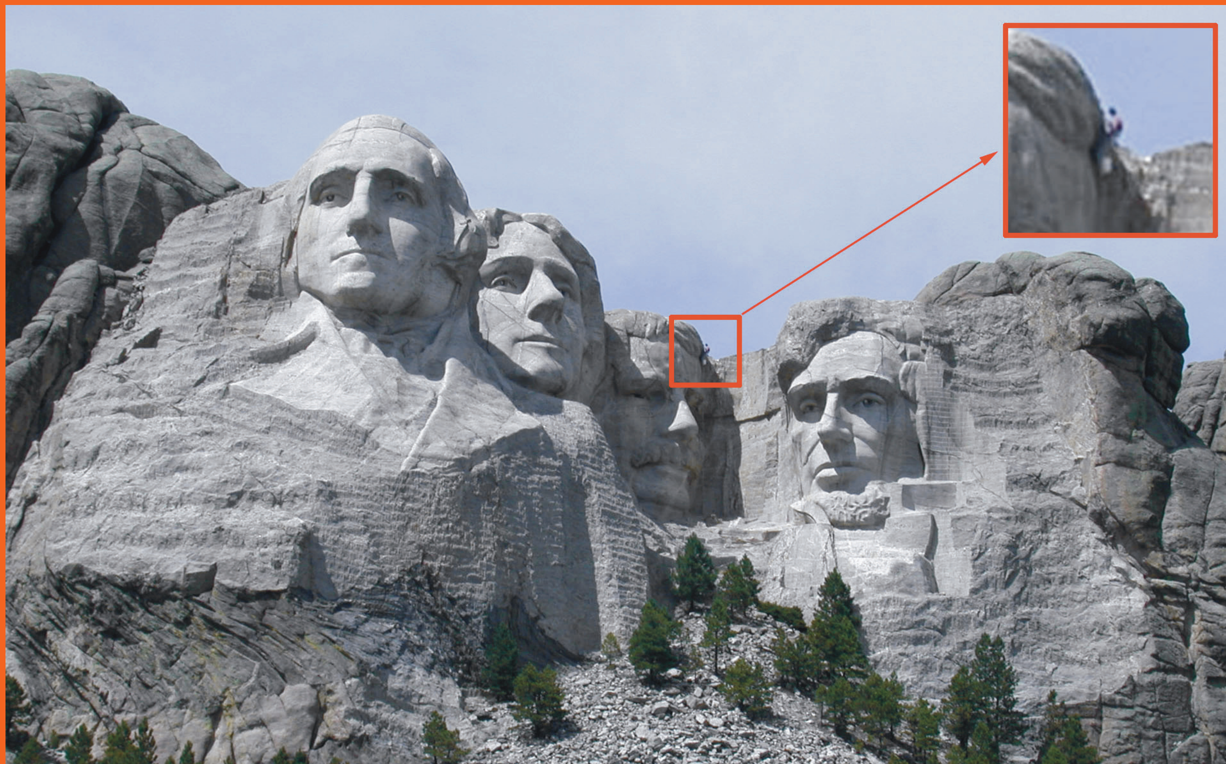

Fig. 1. An anomaly detected on Mount Rushmore.